Evolution of Developmental Regulation in the Vertebrate FgfD Subfamily

RICHARD JOVELIN1, YI-LIN YAN2, XINJUN HE2, JULIAN CATCHEN2,3, ANGEL AMORES2, CRISTIAN CANESTRO2, HAYATO YOKOI2, AND JOHN H. POSTLETHWAIT2

1Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon
2Institute of Neuroscience, University of Oregon, Eugene, Oregon
3Department of Computer and Information Science, University of Oregon, Eugene, Oregon

ABSTRACT Fibroblast growth factors (Fgfs) encode small signaling proteins that help regulate embryo patterning. Fgfs fall into seven families, including FgfD. Nonvertebrate chordates have a single FgfD gene; mammals have three (Fgf8, Fgf17, and Fgf18); and teleosts have six (fgf8a, fgf8b, fgf17, fgf18a, fgf18b, and fgf24). What are the evolutionary processes that led to the structural duplication and functional diversification of FgfD genes during vertebrate phylogeny? To study this question, we investigated conserved syntenies, patterns of gene expression, and the distribution of conserved noncoding elements (CNEs) in FgfD genes of stickleback and zebrafish, and compared them with data from cephalochordates, urochordates, and mammals. Genomic analysis suggests that Fgf8, Fgf17, Fgf18, and Fgf24 arose in two rounds of whole genome duplication at the base of the vertebrate radiation; that fgf8 and fgf18 duplications occurred at the base of the teleost radiation; and that Fgf24 is an ohnolog that was lost in the mammalian lineage. Expression analysis suggests that ancestral subfunctions partitioned between gene duplicates and points to the evolution of novel expression domains. Analysis of CNEs, at least some of which are candidate regulatory elements, suggests that ancestral CNEs partitioned between gene duplicates. These results help explain the evolutionary pathways by which the developmentally important family of FgfD molecules arose and the deduced principles that guided FgfD evolution are likely applicable to the evolution of developmental regulation in many vertebrate multigene families. J. Exp. Zool. (Mol. Dev. Evol.) 312B, 2009. © 2009 Wiley-Liss, Inc.


The development of body form and the differentiation of cells depend on a variety of processes, including cell adhesion accomplished by cell adhesion proteins (e.g. cadherins, integrins, selectins), cell communication performed by extracellular signaling proteins (e.g. Notch, Hedgehog, Wnt, Tgf-β, Fgf), and differential gene expression controlled by transcription factors (e.g. homeobox, bHLH, C2H2-zinc finger, bZIP). Families of structurally related genes encode these proteins and constitute an ancestral toolkit for the construction of animal embryos. Recent genomic investigations reveal a surprisingly complete gene set already present in the last common ancestor of all metazoans (Technau et al., 2005; Guder et al., 2006; Nichols et al., 2006; Kasbauer et al., 2007; Matus et al., 2007), but this ancestral toolkit...
diversified by gene duplication and gene deletion in the great lineages of animal life.

Stem chordates appear to have inherited a subset of toolkit genes present in stem metazoans, and these genes amplified into several members of each subfamily, many in two rounds of whole genome duplication (WGD) (R1 and R2) that appear to have occurred at the base of the vertebrate radiation before the divergence of chondrichthyan and osteichthyan about 530 million years ago (Kumar and Hedges, '98; Kortschak et al., 2001; Ornitz and Itoh, 2001; Lundin et al., 2003; Dehal and Boore, 2005; Garcia-Fernandez, 2005; Bourlat et al., 2006; Delsuc et al., 2006; Guder et al., 2006; Jacob and Lum, 2007; Kitisin et al., 2007; Sundstrom et al., 2008; Yu et al., 2008; Lynch and Wagner, 2009). Genes derived from a WGD event are called "ohnologs" to emphasize their special characters with respect to genes duplicated by other mechanisms, such as tandem gene duplication, unequal crossing-over, or retrotransposition (Wolfe, 2000; Postlethwait, 2007). An example is the family of Hox clusters, which is initially amplified by tandem duplication into a gene cluster that then replicated in vertebrates into four clusters that appear to be related to each other in the relationship ((A,B)/(C,D)), as would be expected by the 2R model (Meulemans and Bronner-Fraser, 2007; Amemiya et al., 2008).

The widespread occurrence of multigene families and subfamilies raises two major questions regarding the evolution of developmental mechanisms: What are the processes by which individual members of gene families evolve their specialized roles in development? And, to what extent does the origin of individual family members by gene duplication contribute to morphological innovation and lineage divergence?

A classic multigene family essential for animal development, the fibroblast growth factor (Fgf) gene family, encodes small signaling proteins deployed in many aspects of embryo patterning (Ornitz and Itoh, 2001). Human and mouse genomes have 22 Fgf genes grouped into seven subfamilies, each containing two to four members (FgfA (1/2), FgfB (3/7/10/22), FgfC (4/5/6), FgfD (8/17/18), FgfE (9/16/20), FgfF (11/12/13/14), and FgfG (19/21/23)) (Itoh and Ornitz, 2004; Popovic et al., 2005). The Fgf gene complement of the urochordate Ciona intestinalis, whose lineage diverged from the vertebrate lineage before the R2 events (Oda et al., 2002; Delsuc et al., 2006; Vienne and Pontarotti, 2006; Wada et al., 2006; Delsuc et al., 2008), helps elucidate the pattern of diversification of mammalian Fgf subfamilies (Satou et al., 2002). Based on strong evidence from sequence similarities, two of the six C. intestinalis Fgf genes are orthologs of FgfD and FgfF subfamilies, respectively, and moderate support, based on conserved protein signature motifs and conserved genomic syntenies, suggests that three other genes are orthologs of the FgfB, FgfC, and FgfE subfamilies (Satou et al., 2002; Popovic et al., 2005). It is not clear whether Ci-FgfL is orthologous to the FgfA subfamily or whether it is unique to C. intestinalis.

Consider the origin of the FgfB, FgfC, and FgfG subfamilies. A representative of FgfG appears to be missing from Ciona and thus either arose by gene duplication in vertebrates or was secondarily lost in the urochordate lineage. Because FGF3, FGF4, and FGF19 of the FgfB, C, and G subfamilies are contiguous in a cluster in human chromosome 11q13.3, these three genes likely arose by tandem duplication. Those tandem duplications must have occurred before R1 because FGF6 and FGF23 (FgfC and FgfG subfamilies) are nearest neighbors in chromosome 12p13.32 (the other FgfB member was apparently lost from this chromosome), and FGF21 and FGF22 (FgfB and FgfG) are linked on human chromosome 19 (Hsa19) (the FgfC member was apparently deleted from this chromosome). This genomic pattern suggests the model that an ancient FgfB/C/G gene tandemly duplicated to FgfB and FgfC/G before the divergence of urochordates and vertebrates and that subsequently, FgfC/G duplicated to FgfC and FgfG before R1; after R2, chromosome rearrangements and gene loss led to the current human genomic arrangement.

Unexpectedly, the cnidarian Nematostella vectensis has at least 13 Fgf genes, twice as many as Ciona and five or six times more than Drosophila melanogaster and Caenorhabditis elegans (Matus et al., 2007). Phylogenetic analysis suggests that nine Nv-Fgfs may be species-specific with no clear relationship to vertebrate Fgfs, whereas four Nv-Fgfs fall within the FgfD subfamily (Matus et al., 2007).

In mammals, the FgfD subfamily includes Fgf8, Fgf17, and Fgf18 (Ornitz and Itoh, 2001) and in zebrafish, fgf24 (Itoh and Ornitz, 2004; Popovic et al., 2005). The vertebrate FgfD phylogeny suggests either the hypothesis that Fgf24 was lost in the tetrapod lineage (Jovelin et al., 2007), or the hypothesis that fgf24 was an innovation in...
teleosts. FgfD members play critical roles in mesoderm and neuronal induction across large phylogenetic distances. For instance Fgf8, Fgf17, and Fgf18 are expressed at the midbrain–hindbrain border (MHB) (Sato et al., 2004). FGF18 is required for skeletal development (Ohuchi et al., 2000; Liu et al., 2002; Ohbayashi et al., 2002), and Fgf8 plays an important role in MHB and limb bud development (Lewandoski et al., 2000; Moon and Capecchi, 2000; Sun et al., 2002; Boulet et al., 2004). Remarkably, FGF signaling in cnidarians may be important during gastrulation and neural induction as well (Matus et al., 2007).

Until recently, the comparison of Fgf gene content between mammals and zebrafish did not reflect the R3 hypothesis, a WGD occurring in the teleost lineage after the teleosts diverged from other ray-fin fish (Sidow, '96; Amores et al., '98; Postlethwait et al., '98; Wittbrodt et al., '98; Meyer and Schartl, '99; Christoffels et al., 2004; Hoegg et al., 2004; Naruse et al., 2004; Meyer and Van de Peer, 2005; Crow et al., 2006). Our previous phylogenetic analysis of the vertebrate FgfD subfamily showed that fgf8 and fgf18 were duplicated early in the teleost lineage and provided evidence for a diversification compatible with the ray-fin fish genome duplication (Jovelin et al., 2007). Duplicates of human fgf6, fgf10, and fgf20 genes have also recently been identified in zebrafish (Itoh and Konishi, 2007).

Because of the central role of FGF signaling in development, mutations in Fgf genes and their receptors cause hereditary human diseases (Itoh, 2007). An understanding of the precise evolutionary relationships among Fgf genes and insights into the mechanisms of their functional diversification are required to fully connect teleost research to human biology. For instance, in mouse, Fgf8 is required for heart development (Abu-Issa et al., 2002; Frank et al., 2002; Ilagan et al., 2006). In zebrafish, fgf8a is required for the expression of cardiac genes (Reifers et al., 2000b), but fgf8b is not expressed in the heart; conversely, in stickleback, fgf8b but not fgf8a is expressed in the heart (Jovelin et al., 2007). After the ray-fin fish genome duplication, independent evolution of regulatory elements in lineages leading to zebrafish and stickleback led to the differential partitioning of fgf8 subfunctions in heart development between paralogs so that orthologous genes in the two species have come to express different functions (Jovelin et al., 2007).

In this study, we investigated how the developmental roles of FgfD subfamily members evolved as these genes originated from R1 to R3. First, we analyzed the conservation of synteny among chordate FgfD genes to discern evolutionary relationships ambiguous in phylogenies (Satou et al., 2002; Itoh and Ornitz, 2004; Popovic et al., 2005; Jovelin et al., 2007). Second, we used in situ hybridization in stickleback and zebrafish, whose lineages diverged early in teleost phylogeny, about 300 million years ago (Hedges, 2002), to examine how functions of duplicate genes evolve in separate lineages. Third, we analyzed the conservation of noncoding regions in the vicinity of FgfD genes in the context of gene phylogeny and gene expression to learn how divergence in regulatory elements in the face of differential gene loss could affect functional divergence of surviving ohnologs.

Results confirmed the presence of an Fgf24 gene at the base of gnathostome vertebrates, the partitioning of functions between Fgf24 and Fgf8, and revealed the origin of teleost fgfD paralogs. Expression analyses identified shared ancestral functions and newly evolved functions in various members of the FgfD group, and exploration of conserved noncoding elements (CNEs) gave strong support for a pattern of functional divergence through subfunctionalization. The general pattern of gene duplication and functional divergence that this study reveals is likely to apply to many other vertebrate multiple gene families.

**MATERIALS AND METHODS**

**Conserved synteny**

Genome sequences were investigated using the Ensembl databases (http://www.ensembl.org/index.html) for zebrafish Zv7, human NCBI 35, mouse NCBI m36, and C. intestinalis version 2.0. Orthologies were determined by best reciprocal BLASTN hit analysis (Wall et al., 2003) automated in a genome-wide scale in our Synteny Database genomic analysis software (Catchen et al., 2009). From the elephant shark genomic database we isolated the following FgfD gene fragments: Fgf8: exon2 AAVX01442442; Fgf17: no fragments identified; Fgf18: exon2 AAVX01015792 and AAVX01460739, exon3 AAVX01088747, and exon4 AAVX01419960; Fgf24: exon2 AAVX01560519 and AAVX01583584, exon3 AAVX01466056 and AAVX01521786, and exon4 AAVX01108229. The absence of Fgf17 is not significant owing to the low (1.4 ×) coverage of the shark genome (Venkatesh et al., 2007).
Gene expression

In situ hybridization probes were made by in vitro transcription of linearized clones and labeled with digoxigenin-UTP. Stickleback probes were synthesized from TOPO/Not-1 linearized clones using T3 RNA polymerase. Zebrafish fgf8a, fgf8b, fgf17, and fgf24 and stickleback fgf8a and fgf8b probes were as described (Reifers et al., 2000a; Draper et al., 2003; Cao et al., 2004; Jovelin et al., 2007). Additional probes were: zebrafish Fgf18, Fgf18+73 GTGTTTGGGTTGGAGT (ENSGACT00000023719) covering most of the coding region; stickleback CCGTCCTGTGGTCT and Fgf18L-620 ATGG-1 Gacfgf18-585 GCTTGCCCCGCTTGCTGACG and Fgf18-578 TTTTTGCTCCGCTTGCTGACT-1

Elements in the reference sequence were masked using human and fugu masks.

RESULTS

Vertebrate FgfD genes originated in two rounds of duplication

To learn the origins of vertebrate FgfD genes, we investigated conserved syntenies among chordate genomes, beginning with nonvertebrate chordates.

The human genome has four segments corresponding to the FgfD region of nonvertebrate chordates:

Sequence similarities between the C. intestinalis FgfD gene Ci-Fgf8/17/18 (NP_001027648, ENSCING00000009302) and vertebrate FgfD members suggest that the FgfD subfamily originated from a single gene already present in stem olfactores, the last common ancestor of urochordates and vertebrates (Satou et al., 2002; Popovic et al., 2005). The R2 hypothesis predicts that four co-orthologs of the prototypic FgfD gene should be present in four paralogous chromosome segments in vertebrate genomes in the absence of chromosomal rearrangements or gene losses. To test this prediction, we examined the conservation of syntenies to compare the genomic neighborhood of the single C. intestinalis Fgf8/17/18 gene with genomic regions surrounding human FgfD genes (Fig. 1). Results showed that many genes flanking Ci-Fgf8/17/18 have human orthologs near FGF8, FGF17, or FGF18. Immediately to the left of Fgf8/17/18 lies ENSCINT00000019802, the ortholog of FBXW4, and immediately to the right of Fgf8/17/18 is XP_002122436, which is the ortholog of human NPM1/2/3, and all three genes are transcribed in the same orientation (Fig. 1A). Human orthologs of these three genes are also contiguous and transcribed in the same orientation on Hsa10 (FBXW4, FGF8, NPM3) (Fig. 1A).

Analysis of conserved noncoding elements

Genomic sequences including the immediate neighbors of each Fgf gene were obtained from Ensembl release 47 (Hubbard et al., 2007) for human and teleosts (zebrafish, and the four percomorph fish: stickleback, medaka, fugu, and green-spotted pufferfish). Some neighbors flanking medaka fgf24 and fugu fgf17 were unavailable owing to short scaffolds, and a flanking neighbor of zebrafish fgf18 was too distant to be included in the analysis. Sequences were aligned using the program LAGAN (Brudno et al., 2003) and CNEs were identified using the VISTA server (Mayor et al., 2000, Frazer et al., 2004). Repeated
suggesting that these may be paralogs from the R1 and R2 genome duplication events.

To determine whether the human orthologs of genes neighboring *C. intestinalis* Fgf8/17/18 were clustered on human chromosomes or widely distributed, we used circle plots. Figure 1B shows that human orthologs of the neighbors of the *C. intestinalis* Fgf8/17/18 gene cluster around the human Fgf8, Fgf17, and Fgf18 genes rather than being splattered over the full extent of each human chromosome. We conclude that syntenies have tended to be conserved in this region of the genome, which would be expected if local inversions were more frequent than translocations during chordate evolution. Furthermore, this result is as predicted from the hypothesis that stem olfactores (urochordates + vertebrates) had a chromosome segment that duplicated twice, giving rise to paralogs on Hsa5, 8, and 10 containing FGF18, FGF17, and FGF8, and another paralog containing parts of Hsa4 or Hsa2p that today lacks an FgD gene.

**The human genome has four FgD-related paralogs**

The hypothesis that FgD genes arose in two rounds of chromosome duplication predicts that a dot plot that localizes paralogs of genes surrounding FGF8 on Hsa10 should identify four paralogs, including parts of Hsa5 and Hsa8, and one other chromosome segment. To test this prediction, we investigated the chromosome locations of the paralogs of all genes in the 10 Mb interval surrounding FGF8, from 98 to 108 Mb on Hsa10. Figure 1C shows dots for genes from the region of Hsa10 surrounding FGF8 and their paralogs on other chromosomes displayed by red pluses in a
column above or below the Hsa10 gene. Results showed that chromosomes 2, 4, 5, and 8 have 13, 13, 8, and 9 paralogs of the genes indicated by gray circles on Hsa10, but all other chromosomes have six or fewer paralogs, with an average of 1.5 paralogs per chromosome. These results support the conclusion from the analysis of conserved syntenies shared between C. intestinalis and human genomes that parts of Hsa2p and/or Hsa4 are members of the FgfD-containing paralogon.

**Teleost FgfD genes reflect R3 and an ohnolog gone missing**

Having identified paralogous chromosome segments in the human genome that have properties expected for “ohnologons” arising from the R1 and R2 WGD events, we turned to understanding the relationship of the six teleost FgfD genes to the one Ciona and three human FgfD genes.

**Is fgf24 an ohnolog gone missing from the human genome?**

Recognizing that the primary fate of duplicate genes is nonfunctionalization (pseudogenation) of one of the two paralogs (Bershtein and Tawfik, 2008), we explored the hypothesis that the predicted fourth R2-derived co-ortholog of the prototypical FgfD gene is Fgf24 and that this gene was preserved in the teleost lineage but was lost in the human lineage, as suggested by phylogenetic analysis (Jovelin et al., 2007). An alternative hypothesis to be ruled out is that fgf24 did not exist in the last common ancestor of teleosts and humans but was a teleost innovation.

The R2 hypothesis makes three predictions. First, the teleost fgf24 gene should be embedded in a chromosome segment that is conserved with the human genome; second, the conserved region should have paralogs near other human FGFD genes; and third, this segment should show conserved synteny with the Ciona Fgf8/17/18 gene neighborhood shown in Figure 1. Here we test these predictions by examining conserved syntenies around zebrafish fgf24, starting at the first gene to its right (Fig. 2).

The gene lying immediately to the right of fgf24 in the zebrafish and stickleback genomes (Q566X7 and ENSGACT00000022099, respectively) encode proteins that form a well-supported sister clade to

![Conserved syntenies for fgf24](image)

*Fig. 2. Conserved syntenies for fgf24. (A) A portion of Hsa2p with orthologs to neighbors of teleost fgf24 genes. (B) A portion of zebrafish (*Danio rerio*) linkage group 14 (Dre14) containing fgf24. (C) A portion of stickleback (*Gasterosteus aculeatus*) linkage group IV (GacIV) containing fgf24. This evidence is as expected if the human orthologs of fgf24 and npm4 were deleted from the human genome. Lines connect orthologs. Positions on chromosomes given in megabases (Mb).*
tetrapod NPM1 and the zebrafish gene npm1; we will call this gene npm4 for now (Fig. 2B and Suppl. Fig. S1). Thus, npm4 is either an npm1 duplicate (e.g. npm1b), which would make fgf24 a duplicate of fgf18, or npm4 is a gene, like its neighbor fgf24, without a human ortholog. The human genome has three NPM genes: NPM1 (5q35), which is adjacent to FGF18 (8p21.3); NPM2, which is the nearest neighbor of FGF17, and NPM3, which is the nearest neighbor of FGF8 (10q24) (Figs. 3B, 4C, 5A). Thus, an FgfD gene and an Npm gene are neighbors for all extant human FgfD genes, as would be expected if this was the ancestral condition before R1, and neither npm4 nor its neighbor fgf24 have unique human orthologs.

The next full gene to the right of fgf24 (ENSDARG00000001273) is annotated as tlx3a but is actually an ortholog of TLX2 (Wotton et al., 2008). Human and zebrafish genomes both have three TLX family genes (Andermann and Weinberg, 2001; Langenau et al., 2002): TLX1 is five genes distant from FGF8 and tlx1 is three genes distant from fgf8a. TLX3 is the second gene from FGF18, whereas tlx3b is 2.8 Mb from fgf18l. TLX2 is in Hsa2p13.1-p12 without a nearby FGFD-family gene, but tlx2 (NP_705937) is the second gene from fgf24, which favors the model that fgf24 is an ohnolog gone missing from the human genome.

The next gene to the right, ENSDARG-00000063176 from Ensembl release 45, is not annotated in Zv7 despite the continued presence of its nucleotide sequence in the position indicated in Figure 2B. ENSDARG00000063176 has as a best human BLAST hit KAZALD1 located at 10q24.31 seven genes from FGF8, but in a phylogenetic tree, it forms, with its ortholog in the pufferfish Tetraodon nigroviridis, an outgroup to the tetrapod zebrafish KAZALD1 clade.

---

**Fig. 3.** Conserved syntenies for Fgf18. (A) A portion of stickleback chromosome GacIV containing fgf18. (B) A portion of zebrafish chromosome Dre10 containing fgf18b. (C) A portion of human chromosome Hsa5 containing FGF18. (D) A portion of Dre14 containing fgf18a. (E) Locations of each region shown related to the whole chromosome. Black genes have orthologs in these sections, gray genes do not.
This phylogenetic grouping suggests that, like npm4 and fgf24, ENSDARG00000063176 is a paralog without a human ortholog and should be called Kazald2 (Wotton et al., 2008). The two neighbors to the immediate left of zebrafish fgf24 (lbx2 and pcgf1, Fig. 2B) have orthologs adjacent on Hsa2p (Fig. 2A). In human, LBX2, PCGF1, and TLX2 are adjacent, but in the zebrafish genome fgf24 and npm4 lie between pcgf1 and tlx2 (Fig. 2A and B). Moving to the right of TLX2, two of the next four genes in human, AUP1 and LOXL3, have orthologs also to the right of fgf24 in zebrafish (Fig. 2A and B). Thus, the zebrafish and human share most of the genes in this region of Hsa2p/Dre14, but orthologs of fgf24 and npm4 are specifically excised from the human genome. Stickleback linkage group IV (GacIV), which has five genes in a row orthologous to the zebrafish segment in Dre14 just discussed, and an ortholog of aup1 syntenic on GacIV, but about half a chromosome away (Fig. 2C). These results suggest that the five genes (lbx2, pcgf1, fgf24, npm4, tlx2) were neighbors before the divergence of zebrafish and stickleback lineages and that inversions in the stickleback lineage removed aup1 from its former neighbors.

The most parsimonious explanation of these data is that fgf24 and npm4 were adjacent in the FgfD paralogon in the last common ancestor of zebrafish.

Fig. 4. Conserved syntenies for Fgf8. (A) A portion of stickleback chromosome GacVI containing fgf8a. (B) A portion of zebrafish chromosome Dre13 containing fgf8b. (C) A portion of human chromosome Hsa10 containing FGF8. (D) A portion of Dre1 containing fgf8b. (E) A portion of GacIX containing fgf8b. Black genes have orthologs in these sections, gray genes do not.
and human, and that they were lost in the human lineage. Note that the orientation of TLX2/tlx2 relative to LBX2/lbx2 and PCGF1/pcgf1 is opposite in human and zebrafish, suggesting that a local inversion stirred the sequences, either in the human or zebrafish lineage. The human FGF18 region can serve as an outgroup to order the vector of evolutionary change: TLX3, NPM1, and FGF18 are all in the same orientation, suggesting that the deletion of the tetrapod Fgf24 and Npm4 ohnologs was associated with a small local inversion.

When did the duplication event that produced Fgf24 occur? BLAST searches of the trace files of the genome sequencing project of the elephant shark Callorhinchus milli revealed that chondrichthyes have at least an Fgf8 gene (exon 2 AAVX01442442) and both an Fgf18 gene (exon 3 AAVX01088747) and an Fgf24 gene (exon 3 AAVX01521786), as shown by exclusive amino acid positions in an alignment of exon-3 of FgfD genes from various vertebrates (Suppl. Fig. S3A). Because the shark genome, which diverged from the bony vertebrates before the divergence of teleost and tetrapod lineages, has an Fgf24 gene, this gene must have already existed in the last common ancestor of all jawed vertebrates. The inclusion of shark sequences for Fgf18 and Fgf24 in a phylogenetic tree of exon 3 results in the topology ((Fgf8,Fgf17)(Fgf18,Fgf24)) as expected from their origin in the R1 and R2 genome duplications (Suppl. Fig. S3B).

Altogether, the best interpretation of these findings is that the vertebrate FgfD subfamily originated from a single gene present in the ancestor of vertebrates and urochordates by two rounds of genome duplication. These R1 and R2 events lead to four Fgf genes in the last common ancestor of teleosts and tetrapods but the ortholog of fgf24 was subsequently lost in the tetrapod lineage.

**fgf18a and fgf18b are duplicates from the teleost genome duplication**

Phylogenetic analysis shows that zebrafish has two genes (fgf18a (NM_001013264) and fgf18b (also called fgf18l; NM_001012379)) in a clade with FGF18 as the most closely related human gene. Zebrafish Fgf18b is sister-taxon to a teleost Fgf18a clade (Itoh and Konishi, 2007; Jovelin et al., 2007; Kikuta et al., 2007) as would be expected from gene duplication after the divergence of ray-fin and lobe-fin fish (Itoh and Konishi, 2007; Jovelin et al., 2007; Kikuta et al., 2007). Four other sequenced teleost genomes have an ortholog of fgf18a but lack an annotated fgf18b and BLAST searches failed to identify an ortholog of fgf18b in genome databases for stickleback, medaka, Tetraodon, or fugu, all of which are percomorph fish. This situation could have arisen if fgf18a and fgf18b arose by the R3 duplication in stem teleosts and fgf18b was lost in the lineage of...
the Percomorpha but was retained in the lineage of the Ostariophysi (including zebrafish). An alternative hypothesis is that fgf18b arose by tandem duplication in the zebrafish lineage after it diverged from the percomorphs.

Conserved syntenies can help distinguish the R3 and tandem duplication hypotheses for the origin of fgf18 paralogs. The fgf18a gene of zebrafish is on Dre14 near fgf24 along with fbxw11, the ortholog of which is separated from FGF18 by a single hypothetical gene (Fig. 3C and D), thus supporting the orthology of zebrafish fgf18a and human FGF18. Note that the C. intestinalis genome has two orthologs of FBXW11, one located 10 genes distant and one 13 genes away from Fgf8/17/18 (NM_001027626 and ENSCINT-00000018903) (Fig. 1A); this is another example of an ancient FgfD-region synteny conserved for 800 million years. A cluster of pcdh1 genes flanking fgf18 (not shown) are, as a group, co-orthologous to PCDH1 protocadherin genes in Hsa5q31, a bit distant but syntenic to FGF18 in Hsa5q35 (Fig. 3C–E). These gene clusters were apparently tandemly duplicated at least partly independently in the zebrafish and human lineages. Together, these conserved syntenies support orthology of fgf18a and FGF18.

The orthologs of several genes located near zebrafish fgf18b on zebrafish LG10 are adjacent to or near FGF18 on Hsa5q (Fig. 3B and C), consistent with the orthology of fgf18b and FGF18 as shown in the phylogeny. The nearest neighbor to fgf18b (Fig. 3B) is an ortholog of FBXW11, which is separated from FGF18 by a single small hypothetical gene (Fig. 3C). Like fgf18a and FGF18 neighborhoods, the fgf18b region of zebrafish and stickleback both have a series of tandemly duplicated pcdh genes. The region surrounding fgf18 in stickleback has at least five neighbors with orthologies to genes in the region surrounding fgf18b in zebrafish, a result expected if the two regions are orthologous (Fig. 3A and B). This is in conflict with the phylogeny, which shows that Fgf18 in stickleback (ENSGACT00000023719) is more closely related to Fgf18a than it is to Fgf18b (Jovelin et al., 2007). If the conserved syntenies correctly suggest orthologies, then the zebrafish Fgf18b sequence is evolving more rapidly than the Fgf18a sequence. Alternatively, gene losses may account for the conflict. If orthologs neighboring stickleback fgf18 were lost near fgf18a but not near fgf18b, then the reciprocal best blast method for identifying orthologies could find only genes near fgf18b and call them as orthologs given the loss of these genes from the neighborhood of fgf18a. Both explanations, however, are consistent with the origin of fgf18a and fgf18b in the R3 event in teleost evolution.

**fgf8 duplicates arose in R3**

Zebrafish has one uncontested ortholog of FGF8 (Brand et al., ’96; Reifers et al., ’98), and a second gene (fgf17a) that was initially assigned as a co-ortholog of tetrapod Fgf17 (Reifers et al., 2000a; Cao et al., 2004), but that was later shown to be a fgf8 duplicate and is now called fgf8b (Itoh and Konishi, 2007; Jovelin et al., 2007; Kikuta et al., 2007). A detailed analysis of conserved syntenies in human, zebrafish, and stickleback supports this conclusion. Orthologs of six genes surrounding human FGF8 in a 2.16 Mb region on chromosome Hsa10q (Fig. 4C) are located in a 0.4 Mb region on zebrafish chromosome Dre1 containing fgf8b (Fig. 4B). The orthologous portion of stickleback linkage group GacIX contains fgf8b and eight genes in the corresponding region of the human genome (Fig. 4A). The orthologs of a somewhat different set of human genes flanking FGF8 lie near fgf8a in zebrafish Dre13 and stickleback LG VI (Fig. 4D and E). We conclude that zebrafish fgf8 (NM_131281) and the so-called fgf17a (NM_182856) are co-orthologs of human FGF8 and support the conclusion that they should be named fgf8a and fgf8b (Itoh and Konishi, 2007; Jovelin et al., 2007; Kikuta et al., 2007). Figure 4 shows that in addition to FGF8 co-orthologs, zebrafish and stickleback have co-orthologs of LBX1 (Q5CZV7 and si:busm1-265n4.1) and C10orf26 (zgc:111946 and Q1ECU4) in these genomic regions, indicating that these duplicated loci also originated in the R3 genome duplication.

**Teleosts have a single copy of fgf17**

Teleost genomes have a single ortholog of FGF17. Several genes in the immediate neighborhood the gene originally called fgf17b on zebrafish Dre8 (Cao et al., 2004) are orthologs of genes that are near FGF17 in Hsa8p21 (Fig. 5A and B). These two segments are largely duplicated between zebrafish and stickleback (Fig. 5B and C). We conclude that fgf17b should be called simply fgf17 (NM_182856) and it is the zebrafish ortholog of FGF17.
Expression patterns of FgfD family genes

With this understanding of the evolutionary relationships of teleost and tetrapod FgfD genes, we wondered how paralog functions evolved as judged by gene expression patterns.

Mid-segmentation stages

*fgf8* paralogs: At mid-segmentation stages (40 hpf stickleback and 16 hpf zebrafish) (Kimmel et al., '95), the expression of *fgf8a* and *fgf8b* are similar in zebrafish and stickleback (Reifers et al., '98, 2000a; Draper et al., 2003; Jovelin et al., 2007), with strong expression of both genes in both species in MBH and somites (Fig. 6A–H). In addition, *fgf8a* is expressed in the dorsal diencephalon and the tailbud in both species. In the heart, stickleback and zebrafish express different *fgf8* paralogs (Jovelin et al., 2007).

*fgf17*: At mid-segmentation in stickleback and zebrafish, *fgf17* is expressed weakly in the central nervous system (CNS), and more strongly and specifically around the tail bud and in the segmental plate and segmented somites (Fig. 6I–L) (Cao et al., 2004; Hamade et al., 2006). The major difference between *sfgf17* (stickleback *fgf17*) and *zfgf17* (zebrafish *fgf17*) is that the zebrafish gene is expressed in the dorsal diencephalon (Fig. 6K). For both species, *fgf8a* and *fgf8b* are both expressed in the MHB, but *fgf17* is not.


*J. Exp. Zool. (Mol. Dev. Evol.)*
and all three genes are expressed in the somites. In the tailbud, fgf8a in both species is expressed in the tailbud itself; fgf8b is not expressed in the tailbud region; and fgf17 is expressed in the segmental mesoderm that surrounds the tailbud in a pattern complementary to that of fgf8a. Both zfgf17 and zfgf8a are expressed in the dorsal diencephalon.

**fgf18:** The expression of zfgf18a and zfgf18b have not been previously described. At mid-segmentation, the single copy of fgf18 in stickleback is expressed weakly in the MHB and in the heart primordium, but not in the somites (Fig. 6M and N). In zebrafish, zfgf18b, but not zfgf18a, is expressed in the MHB like fgf8a, and reciprocally, zfgf18a, but not zfgf18b, is expressed in the heart like zfgf8b (Fig. 6O–R). Although sfgf18 is expressed only weakly if at all in the somites (Fig. 6N), zfgf18a and zfgf18b are both expressed in the somites, but zfgf18a is expressed broadly in somites like zfgf8b (Fig. 6H and P), but zfgf18b is expressed in just a part of each somite like zfgf8a (Fig. 6D and R).

**fgf24:** In mid-segmentation stage stickleback embryos, sfgf24 is expressed in the dorsal diencephalon, MBH, and tailbud, but not in the somites, like sfgf18 (Fig. 6S and T). In zebrafish as in stickleback, fgf24 is expressed in the dorsal diencephalon like fgf8a (Draper et al., 2003). In contrast to sfgf24, zfgf24 is not expressed in the MHB (Fig. 6U and V). In both stickleback and zebrafish, fgf24 is expressed in the unsegmented presomitic mesoderm.

### Long pec stage

**fgf8:** At the “long pec” stage (120 hpf stickleback and 48 hpf zebrafish) (Kimmel et al., ’95), the expression patterns of both fgf8 genes in both species occupy several of the earlier domains, but in addition, in both species, new domains appear for fgf8a (but not fgf8b) in the telencephalon, olfactory epithelium, pharyngeal arches, (Meulemans and Bronner-Fraser, 2007), and apical epidermal ridge (AER) of the pectoral fin bud (Fig. 7A–D).

**fgf17:** At the long pec stage, stickleback sfgf17 is expressed in the retina (Fig. 7E) in a pattern that overlaps in time and space expression of sfgf8a in the eye, (Fig. 7A) and is stronger than sfgf8b expression in the eye at the same stage (Fig. 7C). In addition, sfgf17 and zfgf17 are expressed in the otic vesicle (Fig. 7E and F). sfgf17 was only weakly expressed in the pharyngeal arches compared with zebrafish (Fig. 7E and F). No expression of sfgf17 was apparent in the MHB or in the pectoral fin bud. In zebrafish, zfgf17 is also expressed in the eye, but broadly in several layers. In mouse, the expression patterns of Fgf17 and Fg8 largely overlap, but Fgf17 is expressed in a broader region of the frontal cortex and cerebellar development (Xu et al., 2000; Cholfin and Rubenstein, 2007; Dominguez and Rakic, 2008). The regenerating adult zebrafish heart also expresses fgf17 (Lepilina et al., 2006).

<table>
<thead>
<tr>
<th>Gac 5 dpf</th>
<th>Dre 2 dpf</th>
</tr>
</thead>
<tbody>
<tr>
<td>A dd mhb ov fgf8a</td>
<td>B dd mhb ov fgf8a</td>
</tr>
<tr>
<td>C mhb ov fgf8b</td>
<td>D mhb ov fgf8b</td>
</tr>
<tr>
<td>E ov os ov fgf17</td>
<td>F ov os ov fgf17</td>
</tr>
<tr>
<td>G ov ov ov ov fgf18a</td>
<td>H ov ov ov ov fgf18a</td>
</tr>
<tr>
<td>I ov ov ov ov fgf18b</td>
<td>J ov ov ov ov fgf18b</td>
</tr>
<tr>
<td>K mhb ov ov fgf24</td>
<td>L mhb ov ov fgf24</td>
</tr>
</tbody>
</table>

zebrafish \textit{zfgf18b} only adds expression in the retina and pharyngeal arches (Fig. 7G–I).

\textit{fgf24}: At the long pec stage, stickleback \textit{sfgf24} is expressed in the MHB, olfactory epithelium, most cells of the otic vesicle, endoderm of the pharyngeal pouches, and apical epidermal ridge of the pectoral fin (Fig. 7J). Expression of \textit{sfgf24} in the eye and MHB is broader than the expression of \textit{sfgf8a} and \textit{sfgf18} in these domains (Fig. 7A–G). Here we extend previous descriptions of \textit{zfgf24} expression (Draper et al., 2003; Fischer et al., 2003); \textit{zfgf24} is expressed in a small group of cells in the dorsal MHB, in the olfactory epithelium, endoderm of the pharyngeal pouches, and the otic vesicle. In the AER of the pectoral fin in both stickleback and zebrafish, \textit{fgf24} is expressed broader and stronger than \textit{fgf8a}, with the \textit{fgf8} domain entirely within the \textit{fgf24} domain (Fig. 7A, B, J, K).

**Eyes and ears**

Several \textit{FgfD} genes are expressed in the eyes and ears, but often in different domains. The ganglion cell layer of the eye expresses \textit{sfgf8a}, \textit{zfgf8a}, \textit{sfgf17}, and \textit{sfgf24} (Fig. 8A, C, I, S) (Picker et al., 2005). In the retina, \textit{zfgf24} is expressed in isolated cells scattered in various layers (Fig. 8S, U). Except for the retina and MHB, \textit{zfgf24} and \textit{sfgf24} are expressed similarly.

The developing ear expresses all \textit{FgfD} genes at long pec stages in both stickleback and zebrafish. Some genes are expressed generally throughout
Because some regulatory elements conserved in function are also conserved in sequence, we searched teleost and human FgfD genomic regions for CNEs.

**CNE sharing among FgfD paralogs within species**

*Human:* To discover ancient CNEs derived from the R1 and R2 genome duplication events, we conducted a Vista plot analysis of human FGFD genes. The analysis allowed several conclusions (Fig. 10A). First, CNEs shared between FGF8 and FGF18 were more frequent and longer than those shared by FGF8 and FGF17. Second, CNEs embedded in the genes flanking FGF8 (FBXW4 and NPM3) are conserved with those flanking FGF18, which is located adjacent to NPM1 and near FBXW11. NPM2 is adjacent to FGF17 (Fig. 5A), but CNEs embedded in NPM3 (near FGF8) and NPM1 (near Fgf18) are apparently not conserved with NPM2 (near FGF17). Third, a CNE (labeled “a”) shared by FGF8 and FGF18 is found within intron four of FBXW4—this is significant because an enhancer trap element inserted in a similar position in zebrafish confers an fgf8 expression pattern (Kikuta et al., 2007). Finally, three highly conserved CNEs located 3’ to FGF8 (labeled d, e, and f) are shared by all three human FGFD genes. We conclude that at least three noncoding regions in the genomic neighborhood of FGFD genes have been highly conserved in sequence for at least 500 million years (Kumar and Hedges, ’98; Peterson et al., 2004) since the R1 and R2 genome duplication events. Our results raise the hypothesis that these conserved sequences are regulatory elements that drive the expression of FGFD genes in domains common to all three genes.

*Stickleback:* To test if CNEs near FgfD genes are also conserved in teleosts, we constructed Vista plots for stickleback. Fig. 10B shows conservation for the exons of sfgf8a/sfg8b and scl2a5a/scl2a5b at greater than threshold, but no CNEs shared by the two paralogs. This result is a bit surprising given the similarities of expression patterns between the two sfgf8 ohnologs. In contrast, several CNEs were apparent in the comparison of sfgf8a and sfgf17, although most exons failed to rise above threshold (Minimum Y value on the VISTA plot = 50%, Minimum conservation identity = 70%, Minimum length for a CNE = 100).

Three of the stickleback CNEs (a, b, e) were in introns of fgfD neighbor genes, and two were found

![Fig. 9. Expression of stickleback and zebrafish FgfD genes in the trunk. A, C, E, stickleback embryos. B, D, F, zebrafish embryos. A, B, sfgf18 and sfgf18a. C, D, sfgf18 and sfgf18a. E, F, sfgf24. pan, pancreas; spn, spinal cord neurons.](image-url)
in intergenic region flanking the \textit{fgfD} gene. Note that CNE “a” in Fig. 10B is in the same intron of stickleback \textit{fbxw4} as CNE “a” of human \textit{FBXW4} although it is shared between the \textit{fgf8a} and \textit{fgf17} regions for stickleback and \textit{FGF8} and \textit{FGF18} regions for human. Neither the exons nor any CNEs were detected that were shared between \textit{sfgf8a} and either \textit{sfgf18} or \textit{sfgf24}, suggesting rapid evolution, translocations, or problems with genome assembly. The sparse pattern of CNEs in stickleback \textit{fgfD} genes contrasts with the rather broad CNEs found for human \textit{FGFD} genes.

\textbf{Zebrafish}: For zebrafish (Fig. 10C), a strong CNE (labeled “d” in Fig. 10B and C and “a” in Fig. 10A) was detected in the same intron of \textit{fbxw4} as detected in stickleback and in human, and it was shared by all zebrafish \textit{fgfD} regions except \textit{fgf18b}. An enhancer trap insertion nearby (CLGY667) confers reporter expression in telencephalon, optic stalk, midbrain–hindbrain boundary, somites, heart, olfactory pits, and tail bud in a pattern similar to that of \textit{fgf8} genes (Kikuta et al., 2007). This CNE, conserved since before R1 and maintained by \textit{fgf8a}, \textit{fgf8b}, \textit{fgf17}, \textit{fgf18a}, and \textit{fgf24}, may
play a role in defining expression domains shared by vertebrate FgfD genes.

Despite their similarity in expression patterns (Fig. 8A–H, 9A–D), fgf8a shares fewer CNEs with fgf8b than it does with fgf17, fgf18a, and fgf24 (Fig. 10C). Using fgf8a as base, one CNE (d) is present near five zebrafish FgfD genes, three (b, n, and r) are present near three FgfD genes, eight (f, i, k, l, m, q, w, and y) are near two FgfD genes, and 14 (a, c, e, g, h, j, o, p, s, t, u, v, x, and z) are shared by fgf8a and only one or another FgfD gene. fgf8a shares most CNEs (17) with fgf24 and about an equal number with fgf17 and fgf18a (11 and 10, respectively).

Enhancer trap CLG508 is located in between fgf8a and fbxw4 and confers expression in the apical ectodermal ridge of the pectoral fin bud (Kikuta et al., 2007), but Vista detected no CNE shared by fgf8a and other FgfD genes in this intergenic region despite the expression of both fgf8 and fgf24 in the AER. Enhancer trap CLGY1030 is inserted near CNE “u” in zebrafish and matches the expression pattern of fgf8a only in the tail bud (Kikuta et al., 2007).

CNE sharing among FgfD orthologs

CNEs displayed when comparing FgfD genes within a species show elements conserved from the R2 event, whereas CNEs revealed by comparing a single FgfD gene among species identifies elements conserved since species divergence.

fgf17: Few CNEs appeared in comparisons of the stickleback fgf17 gene to its orthologs in human and several teleosts, although exons were readily apparent (Fig. 11). At least six CNEs were apparent when comparing stickleback and zebrafish fgf17 genes, although the flanking genes were not conserved (Fig. 11). Given that stickleback and pufferfish are more closely related than stickleback and zebrafish, it was surprising to find more sfgf17 CNEs shared with zebrafish than pufferfish. The lack of CNEs near fgf17 in the two pufferfish species could be due in part to incomplete sequencing in this genomic region: fgf17 in fugu (Takifugu rubripes) lies in a small scaffold without predicted neighbors, and numerous sequencing gaps surround fgf17 in the green spotted pufferfish T. nigroviridis. Among the species examined, only stickleback and T. nigroviridis share a neighboring gene (ENSGACG00000003433) (Fig. 11). Close inspection of chromosome segments carrying fgf17 in zebrafish and stickleback shows that global synteny is conserved but gene orders are not, indicating local genomic micro-rearrangements (Fig. 5). Comparison with human reveals rearrangements over still larger genomic distances (Fig. 5). Chromosome breaks near fgf17 may have carried CNEs along with fgf17 neighbors disrupting synteny of coding and noncoding sequences in these regions.

fgf18: BLAST searches identified a single fgf18 gene in stickleback, medaka and both pufferfish and the topology of the fgf18 clade suggests that fgf18b may have been lost in the lineage leading to these species (Jovelin et al., 2007). In contrast to the four percomorph species, zebrafish has two fgf18 genes. A number of sfgf18 CNEs are found associated only with zebrafish fgf18a (b, c, d, e, g, h, i, j, l, n, o, q), a few only with zfgf18b (a, f), and several with both (k, m, r) (Fig. 12). This pattern
of reciprocal distribution of CNEs is what would be expected if the single stickleback fgf18 gene represented much of the ancestral distribution of CNEs and they had become distributed as predicted by subfunctionalization in the zebrafish lineage after gene duplication.

The high density of CNEs surrounding fgf18 suggests that fgf18b may have been lost shortly after the separation of the lineages leading to zebrafish and stickleback and that CNEs may have been maintained by selection in lineages in which only one fgf18 gene subsisted. However, in zebrafish, which retained both fgf18 duplicates, most CNEs found around fgf18 of species retaining just one fgf18 gene have been partitioned between fgf18a and fgf18b (Fig. 12).

fgf24: The high density of CNEs in proximity to fgf24 orthologs in all tested species (Fig. 13) is consistent with our inference that one fgf24 duplicate was lost before the divergence of the lineages leading to zebrafish and stickleback (Jovelin et al., 2007). Two particularly strong

Fig. 12. Most of the CNEs conserved among stickleback, medaka and pufferfish fgf18 are partitioned between fgf18a and fgf18b in zebrafish, consistent with the DDC model (Force et al., '99). The high density of CNEs in teleost species having a single fgf18 gene suggests that the second fgf18 duplicate was lost shortly after the R3 duplication. The region lacking CNEs in tetraodon is owing to a gap in the sequence alignment.

Fig. 13. The high density of CNEs among teleost fgf24 orthologs suggests that the second fgf24 duplicate was lost shortly after the R3 duplication before the partitioning of fgf24 subfunctions. Medaka fgf24 is located on a short scaffold, hence the apparent lack of conservation further away from fgf24.
CNEs (labeled a and d) lie 3' and 5' respectively to \textit{fgf24} and two others (b and c) are within introns of the \textit{fgf24} gene. We checked to see if \textit{Fgfd}-related CNEs could be detected in the location predicted for the missing human \textit{FGF24}, but found none.

\textit{fgf8}: In stickleback and zebrafish, \textit{fgf8} duplicates exhibit common and distinct expression patterns, and comparisons between \textit{fgf8} orthologs show that partitioning of most ancestral functions occurred before the divergence of these species (Jovelin et al., 2007). Thus, we should expect regulatory elements to be largely shared between \textit{fgf8} orthologs in stickleback and zebrafish rather than being species-specific. For comparisons based on stickleback \textit{fgf8a}, at least four CNEs were shared with the human \textit{FGF8} region (Fig. 14A). These included elements in introns of \textit{fbxw4} (CNEs a, b, and d in Fig. 14A) and an element in an intron of \textit{fgf8a} itself (element h). Among the five stickleback \textit{fgf8D} genes, the \textit{fgf8} comparisons were the only ones that revealed any CNEs shared with human. All of the elements \textit{sfgf8a} shared with human were also shared with zebrafish. In addition, zebrafish \textit{fgf8a} shared with stickleback \textit{fgf8a} several elements in the intergenic spaces 3' and 5' to \textit{fgf8a}. As noted for other \textit{fgfD} genes, stickleback \textit{fgf8a} had more and longer CNEs shared with percomorphs than with zebrafish. With few exceptions (see element d, Fig. 14A), the percomorphs had very similar CNE patterns.

The comparison of the stickleback \textit{fgf8b} region to that of human revealed a CNE 3' to the stickleback \textit{fgf8b} gene (element b, Fig. 14B) shared with other teleosts except \textit{T. nigroviridis}. A CNE in an intron of \textit{fgf8b} (element c in Fig. 14B) was

![Fig. 14. Distribution of CNEs among Fgf8 orthologs. CNEs are partitioned in unique sets between teleost \textit{fgf8a} (A) and \textit{fgf8b} (B), reflecting the partitioning of expression domains between \textit{fgf8} duplicates (this work and (Jovelin et al., 2007)). Fewer \textit{fgf8b} CNEs are shared with human \textit{FGF8}.](image-url)
shared with zebrafish and fugu. A large number of CNEs located 5' to fgf8b (e.g. element g) were held in common by several teleosts that were not seen in the comparison with zebrafish.

Therefore, the overall pattern of conservation of noncoding sequences in the proximity of fgf8 correlates with the evolutionary history of fgf8 duplicates and their subsequent functional diversification: fgf8a and fgf8b orthologs have unique sets of CNEs, some of which are shared with human as it would be expected if subfunctions have been partitioned among paralogs.

**DISCUSSION**

**Vertebrate FgfD genes originated as ohnologs in R1 and R2**

To infer the evolutionary history of expression domains and conserved noncoding sequences, it is essential to understand the phylogenetic relationships of the four FgfD genes (Suppl. Fig. S4). The phylogenetic tree of these genes shows a history of (((fgf8, fgf17) fgf18) fgf24) (Jovelin et al., 2007). Several other data sets, however, support the conclusion that the four FgfD genes arose in the R1 and R2 WGD events. Evidence supporting this conclusion comes from four independent sources: (1) The chromosome region surrounding the Ciona FgfD gene contains genes with human orthologs mainly on human chromosome segments containing FGF8, FGF17, and FGF18, as well as part of Hsa2p. (2) An automated search for paralogous chromosome segments around FGF8 revealed chromosome segments surrounding FGF17, FGF18, and a portion of Hsa2p. (3) The elephant shark has a copy of fgf24, demonstrating the origin of this gene before the divergence of teleost and tetrapod genomes. (4) The immediate genomic neighborhood of the teleost fgf24 gene is preserved in the human genome with the rather precise excision of an ortholog to fgf24 and its nearest neighbor npm4. These data support the conclusion that R1 and R2 produced four ohnologs related as (((fgf8, fgf17) (fgf18, fgf24)) and that the tetrapod Fgf24 gene went missing from tetrapod genomes.

**Comparative analysis of expression patterns**

After the R2 genome duplication event, the four FgfD genes would have been preserved either by subfunctionalization (the duplicate preservation by reciprocal partitioning of ancestral gene subfunctions) or neofunctionalization (the origin of novel domains) (Force et al., '99; Hughes, '99; Stoltzfus, '99; Postlethwait et al., 2004; Chain et al., 2008; Conant and Wolfe, 2008), and subsequent to preservation, additional expression domains or other functions could have been lost or “invented”. To analyze these events, we investigated expression patterns of FgfD genes in two teleosts and compared them with published data available on their orthologs in mouse (Suppl. Fig. S4).

The single FgfD gene present in stem olfactores (urochordates+vertebrates) was an ancestor of Fgf8/17/18 in the ascidian urochordate C. intestinalis and its expression pattern serves as an outgroup to the vertebrate FgfD genes. In Ciona, Fgf8/17/18 is expressed in the embryonic CNS coincident with En and Pax2/5/8 (Ikuta and Saiga, 2007). Correspondingly, the vertebrate orthologs of these three Ciona genes (Fg8, Fg17, Fg18, En1, En2, Pax2, Pax5, and Pax8) are co-expressed in the vertebrate CNS at the midbrain–hindbrain boundary (MHB) (Joyner, '96). Although urochordates may lack an MHB organizer secondarily (Canestro et al., 2005; Ikuta and Saiga, 2007), this gene set appears to constitute an ancient regulatory cassette in chordate CNS development. The expression of all four vertebrate FgfD ohnologs in the MHB region in at least one vertebrate species (Suppl. Fig. S4) suggests that this was an ancient expression domain existing before the divergence of urochordates and vertebrates and that this expression domain was retained by all four genes after the R1 and R2 WGD events at the base of the vertebrate radiations.

Besides its expression in the CNS, Fgf8/17/18 is also expressed in C. intestinalis in mesodermal cells lateral and anterior to the CNS expression domain, as well as cells at the tip of the tail (Imai et al., 2002; Ikuta and Saiga, 2007). The relationship of these two urochordate domains to similar domains in the vertebrate somitic and unsegmented mesodermal expression and tail bud domain is uncertain. In the cephalochordate amphioxus, Fgf8/17/18 is expressed transiently in the rostral CNS and in the pharyngeal endoderm, which are likely conserved domains with vertebrates (Muelemans and Bonner-Fraser, 2007).

A substantial number of morphological innovations or elaborations occurred after the divergence of urochordates and vertebrates but before the divergence of bony fishes into ray-fin and lobe-fin clades. These include the elaboration of neural crest and placodes, the invention of bone, and the origin
of paired appendages (Gans and Northcutt, '83; Bassham and Postlethwait, 2005; Jeffery, 2006; Sauka-Spengler and Bronner-Fraser, 2008). Also occurring at about this time period were two rounds of WGD, R1 and R2 (Kortschak et al., 2001; Ornitz and Itoh, 2001; Lundin et al., 2003; Dehal and Boore, 2005; Garcia-Fernandez, 2005; Bourlat et al., 2006; Delsuc et al., 2006; Guder et al., 2006; Jacob and Lum, 2007; Kitisín et al., 2007; Sundström et al., 2008; Yu et al., 2008) that we show here are highly likely to have produced the four $FgfD$ orthologs ($Fgf8$, $Fgf17$, $Fgf18$, and $Fgf24$). Any expression domain that appears in the orthologs of three or four of these four genes in at least one of the three species stickleback, zebrafish, and mouse, is highly likely to have been a site of expression in the last common ancestor of all bony fishes.

In addition to the MHB and somite/unsegmented mesoderm, at least one of the three investigated vertebrates (stickleback, zebrafish, mouse) expresses each $FgfD$ gene in the retina, pharyngeal arches, somites, and otic vesicle (Suppl. Fig. S4). We conclude that these expression domains were probably present in the four orthologous $FgfD$ genes in the last common ancestor of ray-fin and lobe-fin fish and were differentially lost from various genes in different lineages. Because these domains were in common in all four $FgfD$ genes, we infer they were also in common in the unduplicated $FgfD$ gene in the lineage leading to vertebrates before R1 and R2. An alternative hypothesis is that these shared expression domains evolved independently but convergently multiple times, a possibility that seems unlikely based on parsimony. If the hypothesis is true that expression domains shared by all four $FgfD$ genes were present in the last pre-R1-duplication $FgfD$ gene, then that vertebrate ancestor would have had expression domains appropriate for eyes, ears, pharyngeal arches, somites, and MHB. Available data for $FgfD$ expression patterns in cephalochordates and urochordates do not contradict this explanation (Imai et al., 2002; Ikuta and Saiga, 2007); Muelemans and Bonner-Fraser, 2007).

Expression domains in the dorsal diencephalon and the apical epidermal ridge (AER) of the pectoral fin/limb bud are shared by $Fgf8$ and $Fgf24$ (Figs. 6, 7). Furthermore, the examination of mutant phenotypes shows that $Fgf8$ is required for paired pectoral appendage development in tetrapods, which lack an ortholog of $fgf24$, and reciprocally, $fgf24$ is required for paired pectoral appendage development in zebrafish but $fgf8a$ is not (Reifers et al., '98; Lewandoski et al., 2000; Draper et al., 2003) ($fgf8b$ is not expressed in the AER of fish pectoral appendages (Jovelin et al., 2007)). If the R1, R2 evolutionary history of the four $FgfD$ genes is correct, then the pre-duplication $FgfD$ gene should have had an expression domain corresponding to the dorsal diencephalon and AER domains in modern vertebrates. According to this hypothesis, the last ancestor before the vertebrate genome duplications would thus have had an expression domain for the AER, even though, paradoxically, it is unlikely that that organism had paired appendages (Coates, '94). Thus, the ancestral function of this expression domain would have been co-opted into appendage development after paired appendages began to evolve in early vertebrates, and it may have been used in the development of unpaired midline fins (Freitas et al., 2006). Alternatively, one or both of these expression domains could have evolved convergently as neo-functionalizations by the two genes $Fgf8$ and $Fgf24$ well after the R1 and R2 events during the evolution of paired fins.

Some additional expression domains are displayed by just two of the four $FgfD$ genes (Suppl. Fig. 4). Expression in tailbud is shared by $Fgf8$ and $Fgf17$ and in unsegmented mesoderm and pancreas by $Fgf18$ and $Fgf24$. This distribution might be predicted by the ($Fgf8$, $Fgf17$/$Fgf18$, $Fgf24$) lineage hypothesis if these domains had a part of the repertoire of the $Fgf8/17$ and $Fgf18/24$ genes after the R1 WGD event. As mentioned above, the tailbud expression domain may relate to the expression domain at the tip of the tail in Ciona and inherited from stem olfactores (Imai et al., 2002; Ikuta and Saiga, 2007). In addition, both $Fgf8$ and $Fgf18$ share expression in the heart and $Fgf17$ is expressed in zebrafish heart regeneration, consistent with either inheritance from the $FgfD$ gene of stem olfactores followed by loss in the $Fgf24$ clade or, less parsimoniously, independent acquisition by convergent evolution.

Several expression domains appear only on single clades in the tree. These include segmental plate only in teleost $fgf17$ genes, telencephalon only in $Fgf8$ of teleosts and tetrapods, olfactory epithelium only in teleost $fgf8$ genes, and spinal neurons only in $fgf18\sigma$ of teleosts. The most parsimonious explanation of these domains is that they represent neo-functionalization events in the corresponding clades.

**Correlation of CNEs and expression domains**

Because many CNEs are shared by $fgf8\sigma$ and other zebrafish $FgfD$ genes (Fig. 10C), we can
attempt to correlate individual CNEs with expression domains in embryonic zebrafish that are summarized in Supplementary Fig. 4. The first conclusion is that none of the CNEs identified was unique for a single tissue. CNE q is shared only by fgf8a, fgf18a, and fgf24, and only these genes express in the olfactory organ, making this a candidate regulatory element for contributing to olfactory expression, although genes possessing CNE q are expressed in tissues in addition to the olfactory epithelium. The dorsal diencephalon expresses only fgf8a, fgf17, and fgf24, and these genes share CNEs labeled b, d, i, k, l, n, and w. (Fig. 10C); of these shared elements, i, k, l, and w are absent from paralogs that do not show expression in the dorsal diencephalon, so these are candidate CNEs for positive regulatory elements contributing to expression in the dorsal diencephalon of zebrafish embryos. Elements e, g, h, p, t, x, and z are found in both fgf8a and fgf24 and in none of the other FgfD genes, and only these two genes are expressed at 2 dpf, making these elements candidates for controlling expression in the apical ectodermal ridge of the pectoral fin bud. For the retina at 16 hpf, only fgf18b lacks expression, and element d is the only element possessed by all zebrafish FgfD genes except fgf18b, and so element d is a candidate for the early expression in the retina. Other factors may turn on fgf18b in the retina at pharyngula stages. Genes expressed in the heart and the spinal cord neurons have in common the same set of elements (b, c, d, m, n, q, r, s, u, and v). The unsegmented mesenchyme, segmental plate, and pancreas each have unique CNE combinations not shared with other tissues. Because vertebrates can conserve regulatory function without conserving sequence similarity (Fisher et al., 2006), it is therefore likely that CNEs identified around other Fgf genes have a functional role as well.

Examples of subfunction partitioning can be found for a variety of gene duplicates (e.g. Lister et al., 2001; Cresko et al., 2003; Liu et al., 2005). The FgfD subfamily in teleosts provides nevertheless a rare opportunity to test predictions of the DDC model because its members share an origin through the R1, R2, and R3 genome duplication events, and because since R3, different FgfD subfamily members have spent different amounts of time as duplicate copies in various lineages. The fgf24 and fgf17 genes apparently became single copy before the divergence of zebrafish and stickleback lineages; the fgf18a and fgf18b paralogs co-existed from R3 to the present in zebrafish, but became single copy in the stickleback lineage after it diverged from the zebrafish lineage; and the fgf8a and fgf8b paralogs have co-existed from R3 to the present in several lineages. In addition, CNEs identified using phylogenetic footprinting in the proximity of the Fgf8 locus have been shown to have regulatory activity in mouse (Beermann et al., 2006) and in zebrafish (Inoue et al., 2006). Moreover, regulatory modules of fgf8 function extend to intronic sequences of fgf8 neighbors (Kikuta et al., 2007). It is therefore likely that CNEs identified around other Fgf genes have a functional role as well.

The large number of long CNEs detected among the three human FGFD genes contrasts with the fewer shorter CNEs found in zebrafish and stickleback. What could be responsible for this difference? It is not mere antiquity because the last common ancestor of the zebrafish fgf8 and fgf18a genes was in fact the same gene as the last common ancestor of the human FGF8 and FGF18 genes. Thus, either the human lineage preserved more broad CNEs than teleost lineages, perhaps amplifying them locally by tandem duplication, or the zebrafish lineage lost more CNEs after the divergence of human and zebrafish lineages. The pattern of just a few, short, but nevertheless
well-conserved CNEs within stickleback and zebrafish genomes may relate to the additional WGD in the teleost lineage not suffered by the tetrapod lineage (R3). Post-R3 fgfD paralogs may have experienced greater relaxation of selective constraints than post-R2 FgfD genes. After both events, we would expect the partitioning of expression domains by fixation of complementary deleterious mutations in regulatory elements (Force et al., '99; Lynch and Force, 2000).

After R3, one fgf18 duplicate was lost after the divergence of the lineages leading to zebrafish and stickleback resulting in one fgf18 gene in stickleback, medaka, and pufferfish whereas zebrafish retained two fgf18 duplicates. Analysis of CNEs in the proximity of the stickleback fgf18 locus shows extensive conservation between stickleback and other percomorph fish (Fig. 12), suggesting that the integrity of regulatory elements in this genomic region is important for fgf18 function and has been maintained by selection in species having a single fgf18 gene. In contrast, fewer CNEs are shared between zebrafish and the percomorphs, and most CNEs have been partitioned between fgf18a and fgf18b in zebrafish. This pattern of shared and partitioned CNEs makes two predictions. First, the amount and level of conservation among CNEs surrounding fgf18 in percomorphs suggest that little subfunctional partitioning may have occurred before the second fgf18 duplicate was lost in this lineage and thus fgf18 in these species may have retained most of the ancestral pre-R3 fgf18 functions. Second, the pattern of shared CNEs between fgf18a and fgf18b in zebrafish, where both duplicates have been maintained over longer evolutionary time, suggests the partitioning of ancestral pre-R3 fgf18 functions between paralogs. Supporting these predictions, stickleback fgf18 is expressed in the heart field precursor and in the MHB at mid-segmentation, whereas in zebrafish, these expression domains are partitioned between fgf18a and fgf18b, respectively (Fig. 6). At the long pec stage, expression of stickleback fgf18 in the olfactory epithelium and in the MHB is partitioned between zebrafish fgf18a and fgf18b (Fig. 7). Similarly, at the early pharyngula stage, expression in the spinal chord neurons partitioned to zfgf18a (Suppl. Fig. S4). Specifically, one or more of elements “a”, “f”, and “r” in Fig. 12 are candidates for expression of fgf18b in the MHB, and elements “b–e”, “g–j”, and “o–q” are candidates for expression of fgf18a in the nose, spinal neurons, and pancreas, assuming all regulatory elements act positively, although, of course some may act to diminish expression in certain tissues.

Sequenced teleost genomes have a single fgf24 and fgf17 gene, indicating that one duplicate of each locus was lost after R3, before the divergence of the lineages leading to zebrafish and other fish species. The extensive number and length of CNEs around all fgf24 orthologs also supports this hypothesis and suggests that regulatory control elements of fgf24 have been maintained by selection (Fig. 13). By contrast, the lower number of CNEs surrounding Fgf17 genes may be mechanistically related to the multiple local chromosome rearrangements disrupting syntenies of both CNEs and Fgf17 neighboring genes between teleosts and human and within teleosts (Figs. 5, 11). CNEs detached from their target genes by chromosomal rearrangements are expected to be lost by genetic drift (Kikuta et al., 2007). For instance, it is noteworthy that Fgf17 expression is the least conserved between mouse and teleosts among FgfD members (Suppl. Fig. S4). The expression patterns of fgf17 orthologs in stickleback and zebrafish, however, are well conserved, suggesting that important ancestral regulatory elements may have been retained in both lineages, and CNEs “a–e” in Figure 11 are candidates for such elements.

Because one fgf24 duplicate was lost before the divergence of zebrafish and percomorph lineages and one fgf18 duplicate was lost after these lineages diverged, fgf24 duplicates originating from R3 have coexisted over shorter evolutionary time than fgf18 duplicates originating from the same WGD before one copy of each gene was lost. Thus, we would expect the functions of fgf24 orthologs to be more conserved than those of fgf18 orthologs. Indeed, not only does zfgf24 share more and longer CNEs with its orthologs than zfgf18a and zfgf18b share with their orthologs (Figs. 12 and 13) but zfgf24 and sfgf24 expression domains are remarkably conserved with only minor differences in the retina and MHB (Figs. 6.S–V, 7.J–K, 8.S–V), whereas fgf18 orthologs show expression differences in somites, retina and ear (Figs. 6N, P, R, 8M–R).

In summary, this comparison of CNEs and expression domains verifies predictions made by the molecular mechanisms of subfunction partitioning and the DDC model (Force et al., '99) for all teleost fgfD members for which synteny has not been disrupted by local chromosomal rearrangements. Additionally, these results identify a clear relationship, predicted by the subfunctionalization
model, between the length of time that paralogs coexist and their distribution of both CNEs and expression domains. Importantly, this work identifies a number of CNEs that are candidates for regulatory elements for specific expression domains in development. The testing of these domains for developmental functions is an ongoing project.

ACKNOWLEDGMENTS

We thank Dong Liu for discussions on ear morphology, Amanda Rapp, Tim Mason and the University of Oregon Zebrafish Facility for providing animals and excellent fish care, and Poh Kheng Loi and Amber Selix of the Histology Facility for sectioning. This work was supported by the National Center for Research Resources (5R01RR020833) and National Institutes of Health (P01 HD22486); the contents of this study are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH. RJ is an Associate of the IGERT Training Program for Development, Evolution, and Genomics, NSF DGE-9972830.

LITERATURE CITED


Dichmann DS, Miller CP, Jensen J, Scott Heller R, Serup P. 2003. Expression and misexpression of members of the FGF


J. Exp. Zool. (Mol. Dev. Evol.)


